Ministry of Economic Development Manatū Ōhanga
Intellectual Property Office of New Zealand

PCT/NZ03/00294

CERTIFICATE

This certificate is issued in support of an application for Patent registration in a country outside New Zealand pursuant to the Patents Act 1953 and the Regulations thereunder.

I hereby certify that annexed is a true copy of the Provisional Specification as filed on 24 December 2002 with an application for Letters Patent number 523384 made by THE HORTICULTURE AND FOOD RESEARCH INSTITUTE OF NEW ZEALAND LIMITED.

Dated 9 January 2004.

Neville Harris

Neville Harris
Commissioner of Patents, Trade Marks and Designs

COMMISSIONER OF PATENTS, TRADEMARKS & DESIGNS
NEW ZEALAND

330 High Street, Lower Hutt and 17 Toop Street, Seaview
PO Box 30687, Lower Hutt, New Zealand or DX SX 11129, Wellington
Phone: +64 4 560 1600 Fax: +64 4 560 1691 www.ipono.govt.nz

NEW ZEALAND PATENTS ACT 1953

PROVISIONAL SPECIFICATION

ENZYMES AND POLYNUCLEOTIDES ENCODING THE SAME

We, THE HORTICULTURE AND FOOD RESEARCH INSTITUTE OF NEW ZEALAND LIMITED, a New Zealand company and Crown Research Institute (under the Crown Research Institutes Act 1992), having a place of business at Mt Albert Research Centre, 120 Mt. Albert Road, Mt Albert, New Zealand New Zealand do hereby declare this invention to be described in the following statement:

ENZYMES AND POLYNUCLEOTIDES ENCODING THE SAME

Technical Field

The present invention relates to the enzyme germacrene-D/*delta*-cadinene synthase and to polynucleotide sequences encoding the enzyme. The invention also relates to nucleic acid constructs, vectors and host cells incorporating the polynucleotide sequences. It further relates to the production of a germacrene-D/*delta*-cadinene synthase and more specifically to the production of germacrene-D and *delta*-cadinene and their use in products such as an insect attractant/deterrent, anti-bacterial and anti-fungal agents, flavours and fragrances and other products. Germacrene-D and *delta*-cadinene may also be used to produce further products with characteristic aromas useful as flavours and fragrances, or products that are useful as insect attractants/deterrents or antimicrobial agents.

Background Art

Germacrene-D (Figure 1) is a cyclic sesquiterpene hydrocarbon ($C_{15}H_{24}$; 7-iso-propyl-10-methyl-4-methylene-cyclodeca-5,10-diene) that is either constitutively present or induced in a wide range of plant species. *Delta*-Cadinene (Figure 2) is a cyclic sesquiterpene hydrocarbon that is either constitutively present or induced in a wide range of plant species.

The biosynthetic pathway for the sesquiterpenes branches off from the general terpenoid pathway, beginning with the allylic diphosphate ester farnesyl diphosphate (FDP, also called FPP) (Bohlmann, *et al.*, Proc. Natl. Acad. Sci. U. S. A. 95, 4126-4133 (1998), Cane and Bowser, Bioorg. Med. Chem. Lett. 9, 1127-1132 (1999), Davis and Croteau, Top. Curr. Chem 209, 53-95 (2000)). Germacrene-D and *delta*-Cadinene are both synthesised from FDP in a reaction that proceeds through a carbocation intermediate (Figure 3) and are catalysed by the sesquiterpene synthases germacrene-D synthase

(Guterman *et al* Plant Cell 14, 2325-2338 (2002)) and *delta*-cadinene synthase (Benedict *et al* Plant Phys 125, 1754-1765 (2001)) respectively. The pathway for sesquiterpene biosynthesis, the acetate/mevalonate pathway, is localised to the cytoplasm; in contrast to the pathways for monoterpene and diterpene biosynthesis, which occur in the chloroplast (Lange, *et al.*, Proc. Natl. Acad. Sci. U. S. A 97, 13172-13177 (2000)).

All known plant terpene synthases, however, whether monoterpene, sesquiterpene or diterpene, appear to be closely related. Similarities include the positioning of intron sequences (Trapp and Croteau, Genetics 158, 811-832 (2001)) and the presence of conserved sequences, such as an aspartate-rich DDXX(D,E) motif (Lesburg, *et al.*, 8, 695-703 (1998)). This motif is involved in the binding of metal ions, usually Mg^{2+}, that are necessary for catalysis.

Germacrene-D is considered to be a key intermediate in the biosynthesis of many sesquiterpenes (Yoshihara et al., Tetrahed Lett 2263-2364 (1969); Bülow and König,Phytochem 55, 141-168 (2000)). Furthermore, it has been shown to increase attraction of and oviposition by the tobacco budworm moth *Heliothis virescens* (Mozuraitis *et al* Chem Senses 27, 505-509 (2002)) and is also a sex stimulant for the male American cockroach (*Periplanata americana* L.) (Nishino *et al.*, J Insect Physiol 23, 415-419 (1977). Germacrene-D, as a component of certain essential oils, has also been shown to possess antibacterial properties (Juteau *et al* Fitoterapia 73, 532-535 (2002)). To date a gene for germacrene-D synthase has been isolated only from rose (Guterman *et al* Plant Cell 14, 2325-2338 (2002)), which has been shown to produce germacrene-D only.

Cadinene is the first intermediate in the conversion by *delta*-cadinene synthase of FDP to sesquiterpene phytoalexins in cotton (*Gossypium barbadense*) (Benedict *et al*Plant Phys 125, 1754-1765 (2001); Davis and Essenberg, Phytochem 39, 553-567 (1995); Davis *et al.*, Phytochem 41, 1047-1055 (1996)), and is the precursor of desoxyhemigossypol and hemigossypol defense sesquiterpenes. To date four cDNAs for *delta*-cadinene synthase have been isolated from *Gossypium arboreum* and characterisation of at least one of these

has been reported (Chen et al, Arch. Biochem Biophys 324, 255-266 (1995); Meng *et al*, J. Nat. Prod. 62, 248-252 (1999))

The applicants have identified a polynucleotide encoding a multifunctional germacrene-D/delta-cadinene synthase which facilitates the production of both germacrene-D and delta-cadinene together in biofermentation processes. The polynucleotide can also be used to co-ordinately manipulate production of both germacrene-D and delta-cadinene in transgenic plants to alter fragrance/flavour characteristics and/or plant pathogen interactions. The gene can also be used as a marker in marker assisted breeding to discover plant material with altered germacrene-D and delta-cadinene composition.

Summary of the Invention

In a first aspect the invention provides an isolated polynucleotide encoding germacrene-D/*delta*-cadinene synthase.

In a further aspect the invention provides an isolated polynucleotide having the sequence shown in Figure 4 or a fragment or variant thereof encoding a polypeptide with germacrene-D/*delta*-cadinene synthase activity.

The polypeptides of the invention are multifunctional. They are capable of facilitating the conversion of FDP to a mixture of germacrene-D and *delta*-cadinene and other sesquiterpenes.

In another aspect the polypeptides of the invention are capable of facilitating the conversion of FDP to a mixture of germacrene-D and *delta*-cadinene in a ratio of 5:95 to 95:5, preferably 10:90 to 90:10, more preferably 70:30 to 90:100.

In a further aspect, the invention provides an isolated polynucleotide encoding the polypeptide shown in Figure 5 or encoding a variant or a fragment of that sequence which has germacrene-D/*delta*-cadinene synthase activity.

In a further aspect the invention provides an isolated germacrene-D/*delta*-cadinene synthase polypeptide.

In yet a further aspect, the invention provides an isolated germacrene-D/*delta*-cadinene synthase having the sequence shown in Figure 5 or a fragment or variant thereof with germacrene-D/*delta*-cadinene synthase activity.

The polypeptides of the invention are useful for *in vitro* preparation of germacrene-D and/or *delta*-cadinene.

In a further aspect the invention provides a vector comprising a polynucleotide of the invention.

In yet a further aspect the invention provides a genetic construct comprising in the 5'-3' direction

(a) a promoter sequence; and

(b) an open reading frame polynucleotide encoding a polypeptide of the invention

Preferably the genetic construct also comprises a termination sequence.

In another aspect the invention provides a genetic construct comprising in the 5'-3' direction

(a) a promoter sequence; and

(b) a polynucleotide which hybridizes to a polynucleotide encoding a polypeptide of the invention

Preferably the genetic construct also comprises a termination sequence.

In a further aspect the invention provides a host cell comprising a genetic construct of the invention.

In still a further aspect, the invention provides a transgenic plant cell which includes a genetic construct of the invention.

In a yet further aspect, the invention provides a plant cell which has been modified to alter expression of a germacrene-D/*delta*-cadinene synthase.

In addition the invention provides a transgenic plant comprising such cells.

In another aspect the invention provides a method for preparing germacrene-D and/or *delta*-cadinene comprising the steps of

(a) culturing a cell which has been genetically modified with a polynucleotide of the invention to provide increased germacrene-D/*delta*-cadinene synthase activity;

(b) providing the cell with farnesyl diphosphate (FDP) or geranyl diphosphate (GDP) if necessary; and

(c) separating the germacrene-D and/or *delta*-cadinene produced.

This method of the invention allows use of biofermentation for a convenient method for preparing the product.

In further aspect the invention provides a method for altering germacrene-D and/or *delta*-cadinene production in a transgenic plant comprising the steps of

(a) introducing into a plant cell a vector comprising a polynucleotide of the invention;

(b) cultivating the transgenic cell under conditions conducive to regeneration and mature plant growth

Brief Description of Drawings

The present invention will be better understood with reference to the accompanying drawings in which:

Figure 1 shows the structures of the enantiomers of germacrene-D.

Figure 2 shows the structure of *delta*-cadinene.

Figure 3 shows the pathway for the synthesis of germacrene-D and *delta*-cadinene from FDP.

Figure 4 shows the polynucleotide sequence that encodes germacrene-D/*delta*-cadinene synthase. The sequence was obtained from a cDNA library that was constructed using mRNA extracted from *Actinidia deliciosa* petals.

Figure 5 shows the predicted amino acid sequence of germacrene-D/*delta*-cadinene synthase from *A. deliciosa* petals.

Figure 6 shows a GC-MS trace of headspace above bacterial cultures harbouring a pET30a control vector (a) or pET30a/75565 (b). FDP was added to both cultures.

Figure 7 shows a GC-MS trace of headspace above partially purified cell free extracts (in binding buffer) from bacterial cultures harbouring a pET30a control vector (a) or pET30a/75565 (b). FDP was added to both cultures.

Detailed Description

In one embodiment of the invention, cells genetically modified to exhibit germacrene-D/*delta*-cadinene synthase activity are used for the production of germacrene-D and/or *delta*-cadinene. While the cells may potentially be of any cell type that can be grown in culture, it is currently preferred to use bacteria or yeast cells for producing germacrene-D and/or *delta*-cadinene (and its oxidation products or derivatives). Preferred cells for use in the biofermentation processes of this embodiment are GRAS microbes; eg *E. coli, Lactobacillus* spp and other non-pathogenic bacteria or yeasts such as brewers yeast.

Germacrene-D and its co- or by- products may be used as flavour and fragrance additives, an antimicrobial agent, as a pheromone for attracting or repelling insects. Delta-cadinene and its co-or by- products may be used as flavour and fragrance additives (eg baking goods, detergents, cosmetics, chewing gum), and as a pheromone repellent, particularly for cotton. Both compounds may be used as precursors for synthesis of other compounds of value.

In another aspect of the invention, the polynucleotides of the invention are used to prepare transgenic plants that over-express the germacrene-D/*delta*-cadinene synthase in at least some parts of the plant. In this way the invention is used to impart fragrance to flowers, repel or attract insects (either as indicator plants, host plants, or alternative hosts) or impart an altered flavour to fruit or prevent disease in fruit, or to extract pharmaceutical products or animal or insect efficacious extracts.

In one particular aspect the polynucleotides of the invention are used in plants of the order Actinidiaceae, particularly in the genus *Actinidia* to provide increased fragrance in flowers.

In another aspect polynucleotides of the invention are used to decrease germacrene-D/*delta*-cadinene synthase activity in kiwi fruit. This may be achieved in several ways, for example by genetically modifying the kiwi fruit so that germacrene-D/*delta*-cadinene

synthase polynucleotide is transcribed in an antisense orientation which results in decreased germacrene-D/*delta*-cadinene synthase translation. Such fruit may then be more resistant to certain diseases and pests.

Altering the levels of germacrene-D and/or *delta*-cadinene in flowers may will alter attraction/repulsion of pollinators also affect invasion by pathogens.

In another aspect the invention provides a method useful in kiwifruit breeding. Segments of the polynucleotide sequences of the invention may be used as probes to investigate the genetic makeup of candidate kiwi fruit varieties with respect to germacrene-D/*delta*-cadinene synthase activity. The presence of high levels of polynucleotides encoding germacrene-D/*delta*-cadinene synthase activity in the flowers of kiwifruit may be used to identify kiwifruit with altered fragrance.

The amino acid sequence of one polypeptide, a germacrene-D/*delta*-cadinene synthase from kiwifruit, and that of the polynucleotide sequence encoding it are given in Figures 5 and 4 respectively. It will however be appreciated that the invention is not restricted only to the polynucleotide/polypeptide having the specific nucleotide/amino acid sequence given in Figures 4 and 5. Instead, the invention also extends to variants of the polynucleotide/polypeptide of Figures 4 and 5 which possess or encode germacrene-D/*delta*-cadinene synthase activity.

The term "polynucleotide(s)" as used herein means a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases and includes DNA and corresponding RNA molecules, including hnRNA and mRNA molecules, both sense and anti-sense strands, and comprehends cDNA, genomic DNA and recombinant DNA, as well as wholly or partially synthesized polynucleotides. An hnRNA molecule contains introns and corresponds to a DNA molecule in a generally one-to-one manner. An mRNA molecule corresponds to an hnRNA and DNA molecule from which the introns have been excised. A polynucleotide may consist of an entire gene, or any portion thereof. Operable anti-sense polynucleotides may comprise a fragment of the corresponding

polynucleotide, and the definition of "polynucleotide" therefore includes all such operable anti-sense fragments.

The term 'polypeptide(s)' as used herein includes peptides, polypeptides and proteins.

The phrase "variants with germacrene-D/*delta*-cadinene synthase activity" is used in recognition that it is possible to vary the amino acid/nucleotide sequence of a polypeptide/polynucleotide while retaining substantially equivalent functionality. The equivalent can be, for example, a fragment of the polypeptide, a fusion of the polypeptide with another polypeptide or carrier, or a fusion of a fragment with additional amino acids.

An "isolated" polypeptide is a polypeptide that has been identified and separated or recovered to be largely free of components of its natural environment, (that is so that the polypeptide comprises at least 50% of the polypeptides from its natural environment, preferably at least 80%, more preferably at least 90%). The term "isolated" polypeptide includes polypeptides in situ within recombinant cells. However generally isolated polypeptides will be prepared by at least one purification step.

An "isolated" polynucleotide is a nucleotide molecule that is identified and separated from at least one contaminant polynucleotide with which it is ordinarily associated.

Variant polynucleotide sequences also include equivalent sequences, which vary in size, composition, position and number of introns, as well as size and composition of untranslated terminal regions. Variant polynucleotides also include those encoding functionally equivalent polypeptides.

It will be understood that a variety of substitutions of amino acids is possible while preserving the structure responsible for activity of the polypeptides. Conservative substitutions are described in the patent literature, as for example, in United States Patent No 5,264,558 or 5,487,983. It is thus expected, for example, that interchange among non-polar aliphatic neutral amino acids, glycine, alanine, proline, valine and isoleucine,

would be possible. Likewise, substitutions among the polar aliphatic neutral amino acids, serine, threonine, methionine, asparagine and glutamine could be made. Substitutions among the charged acidic amino acids, aspartic acid and glutamic acid, could probably be made, as could substitutions among the charged basic amino acids, lysine and arginine. Substitutions among the aromatic amino acids, including phenylalanine, histidine, tryptophan and tyrosine are also possible. Such substitutions and interchanges are well known to those skilled in the art.

Equally, nucleotide sequences encoding a particular product can vary significantly simply due to the degeneracy of the nucleic acid code.

A polynucleotide or polypeptide sequence may be aligned, and the percentage of identical nucleotides in a specified region may be determined against another sequence, using computer algorithms that are publicly available. An exemplary algorithm for aligning and identifying the similarity of polynucleotide sequences is the BLASTN algorithm. The similarity of polypeptide sequences may be examined using the BLASTP algorithm. Both the BLASTN and BLASTP software are available on the NCBI anonymous FTP server (ftp://ncbi.nlm.nih.gov) under /blast/executables/. The BLASTN algorithm version 2.0.4 [Feb-24-1998], set to the default parameters described in the documentation of variants according to the present invention. The use of the BLAST family of algorithms, including BLASTN and BLASTP, is described at NCBI's website at URL http://www.ncbi.nlm.nih.gov/BLAST/newblast.html and in the publication of Altschul *et al.*, Nucleic Acids Res. 25, 3389-34023 (1997).

The following running parameters are preferred for determination of alignments and similarities using BLASTN that contribute to E values (as discussed below) and percentage identity: Unix running command: blastall -p blastn -d embldb -e 10 -G 1 -E 1 -r 2 -v 50 -b 50 -I queryseq -o results; and parameter default values:
-p Program Name [String]
-d Database [String]
-e Expectation value (E) [Real]

-G Cost to open a gap (zero invokes default behaviour) [Integer]
-E Cost to extend a cap (zero invokes default behaviour) [Integer]
-r Reward for a nucleotide match (blastn only) [Integer]
-v Number of one-line descriptions (V) [Integer]
-b Number of alignments to show (B) [Integer]
-i Query File [File In]
-o BLAST report Output File [File Out] Optional
For BLASTP the following running parameters are preferred: blastall -p blastp -d swissprotdb -e 10 -G 1 -E 1 -v 50 -b 50 -I queryseq -o results
-p Program Name [String]
-d Database [String]
-e Expectation value (E) [Real]
-G Cost to open a gap (zero invokes default behaviour) [Integer]
-E Cost to extend a cap (zero invokes default behaviour) [Integer]
-v Number of one-line descriptions (v) [Integer]
-b Number of alignments to show (b) [Integer]
-i Query File [File In]
-o BLAST report Output File [File Out] Optional

The "hits" to one or more database sequences by a queried sequence produced by BLASTN, BLASTP, or a similar algorithm, align and identify similar portions of sequences. The hits are arranged in order of the degree of similarity and the length of sequence overlap. Hits to a database sequence generally represent an overlap over only a fraction of the sequence length of the queried sequence.

The BLASTN algorithm also produces "Expect" or E values for alignments. The E value indicates the number of hits one can "expect" to see over a certain number of contiguous sequences by chance when searching a database of a certain size. The Expect value is used as a significance threshold for determining whether the hit to a database, such as the preferred EMBL database, indicates true similarity. For example, an E value of 0.1 assigned to a hit is interpreted as meaning that in a database of the size of the EMBL

database, one might expect to see 0.1 matches over the aligned portion of the sequence with a similar score simply by chance. By this criterion, the aligned and matched portions of the sequences then have a 90% probability of being the same. For sequences having an E value of 0.01 or less over aligned and matched portions, the probability of finding a match by chance in the EMBL database is 1% or less using the BLASTN algorithm.

According to one embodiment, "variant" polynucleotides, with reference to each of the polynucleotides of the present invention, preferably comprise sequences having the same number or fewer nucleic acids than each of the polynucleotides of the present invention and producing an E value of 0.01 or less when compared to the polynucleotide of the present invention. That is, a variant polynucleotide is any sequence that has at least a 99% probability of being the same as the polynucleotide of the present invention, measured as having an E value of 0.01 or less using the BLASTN algorithm set at the parameters discussed above.

Variant polynucleotide sequences will generally hybridize to the recited polynucleotide sequence under stringent conditions. As used herein, "stringent conditions" refers to prewashing in a solution of 6X SSC, 0.2% SDS; hybridizing at 65°C, 6X SSC, 0.2% SDS overnight; followed by two washes of 30 minutes each in 1X SSC, 0.1% SDS at 65°C and two washes of 30 minutes each in 0.2X SSC, 0.1% SDS at 65°C. The variant polynucleotide sequences of the invention are at least 50 nucleotides in length.

Variant polynucleotides also include sequences which have a sequence identity of at least 60%, generally 70%, preferably 80%, more preferably 90%, even more preferably 95%, very preferably 98% and most preferably 99% or more to the nucleotide sequence given in the sequence listing herein.

In general, polypeptide sequences that code for the germacrene-D/*delta*-cadinene synthases of the invention will be at least 60%, preferably 70%, and even 80%, 85%, 90%, 95%, 98%, most preferably 99% homologous or more with the disclosed amino

acid sequence. That is, the sequence similarity may range from 60% to 99% or more. In addition the invention includes polynucleotide sequences encoding these amino acid sequences.

Also encompassed by the invention are fragments of the polynucleotide and polypeptide sequences of the invention. Polynucleotide fragments may encode protein fragments which retain the biological activity of the native protein. Alternatively, fragments used as hybridisation probes generally do not encode biologically active sequences. Fragments of a polynucleotide may range from at least 15, 20, 30, 50, 100, 200, 400 or 1000 contiguous nucleotides up to the full length of the native polynucleotide sequences disclosed herein.

Fragments of the polypeptides of the invention will comprise at least 5, 10, 15, 30, 50, 75, 100, 150, 200, 400 or 500 contiguous amino acids, or up to the total number of amino acids in the full length polypeptides of the invention.

Variant is also intended to allow for rearrangement, shifting or swapping of one or more nucleotides or domains/motifs (from coding, non-coding or intron regions) from genes (including terpene synthases) from the same or other species, where such variants still provide a functionally equivalent protein or polypeptide of the invention or fragment thereof.

It is, of course, expressly contemplated that homologs to the specifically described germacrene-D/*delta*-cadinene synthase having the sequence of Figure 5 exist in other plants. Such homologs are also "variants" as the phrase is used herein.

A polynucleotide sequence of the invention may further comprise one or more additional sequences encoding one or more additional polypeptides, or fragments thereof, so as to encode a fusion protein. Systems for such recombinant expression include, but are not limited to, mammalian, yeast, bacteria and insect systems.

DNA sequences from plants other than *A. deliciosa* which are homologs of the germacrene-D/*delta*-cadinene synthase of Figure 4 may be identified (by computer-aided database searching) and isolated following high throughput sequencing of cDNA libraries prepared from such plants. Alternatively, oligonucleotide probes based on the sequence of Figure 5 can be synthesized and used to identify positive clones in either cDNA or genomic DNA libraries from other plants by means of hybridization or PCR techniques. Probes should be at least about 10, preferably at least about 15 and most preferably at least about 20 nucleotides in length. Hybridization and PCR techniques suitable for use with such oligonucleotide probes are well known in the art. Positive clones may be analyzed by restriction enzyme digestion, DNA sequencing or the like.

The polynucleotides of the present invention may be generated by synthetic means using techniques well known in the art. Equipment for automated synthesis of oligonucleotides is commercially available from suppliers such as Perkin Elmer/Applied Biosystems Division (Foster City, CA) and may be operated according to the manufacturer's instructions.

As a result of the identification of the polypeptides and polynucleotides of the invention germacrene-D/*delta*-cadinene synthase activity may be modulated in plants. Modulation may involve a reduction in the expression and/or activity (i.e. silencing) of the polypeptide.

Any conventional technique for effecting this can be employed. Intervention can occur post-transcriptionally or pre-transcriptionally. Further, intervention can be focused upon the gene itself or on regulatory elements associated with the gene and which have an effect on expression of the encoded polypeptide. "Regulatory elements" is used here in the widest possible sense and includes other genes which interact with the gene of interest.

Pre-transcriptional intervention can involve mutation of the gene itself or of its regulatory elements. Such mutations can be point mutations, frameshift mutations, insertion

mutations or deletion mutations. These latter mutations include so called "knock-out" mutations in which expression of the gene is entirely ablated.

Examples of post-transcriptional interventions include co-suppression or anti-sense strategies, a dominant negative approach, or techniques which involve ribozymes to digest, or otherwise be lethal to, RNA post-transcription of the target gene. Co-suppression can be effected in a manner similar to that discussed, for example, by Napoli *et al. Plant Cell 2, 279-290* (1990) and de Carvalho Niebel *et al. Plant Cell 7, 347-358* (1995). In some cases, it can involve over-expression of the gene of interest through use of a constitutive promoter. It can also involve transformation of a plant with a non-coding region of the gene, such as an intron from the gene or 5' or 3' untranslated region (UTR).

Anti-sense strategies involve expression or transcription of an expression/transcription product capable of interfering with translation of mRNA transcribed from the target gene. This will normally be through the expression/transcription product hybridising to and forming a duplex with the target mRNA.

The expression/transcription product can be a relatively small molecule and still be capable of disrupting translation from the mRNA. However, the same result is achieved by expressing the whole polynucleotide in an anti-sense orientation such that the RNA produced by transcription of the anti-sense oriented gene is complementary to all or part of the endogenous target mRNA.

Anti-sense strategies are described generally by Robinson-Benion *et al.* Methods in Enzymol 254, 363-375 (1995) and Kawasaki *et al.*, Artific. Organs 20, 836-845 (1996).

Genetic constructs designed for gene silencing may include an inverted repeat. An 'inverted repeat' is a sequence that is repeated where the second half of the repeat is in the complementary strand, e.g.,
5'-GATCTA.....TAGATC-3'

3'-CTAGAT.....ATCTAG-5'

The transcript formed may undergo complementary base pairing to form a hairpin structure provided there is a spacer of at least 3-5 bp between the repeated regions.

Another approach is to develop a small antisense RNA targeted to the transcript equivalent to an miRNA (Llave *et al.*, Science 297, 2053-2056 (2002) that could be used to target gene silencing.

The ribozyme approach to regulation of polypeptide expression involves inserting appropriate sequences or subsequences (eg. DNA or RNA) in ribozyme constructs (McIntyre Transgenic Res. 5 257-262 (1996)). Ribozymes are synthetic RNA molecules that comprise a hybridizing region complementary to two regions, each of which comprises at least 5 contiguous nucleotides of a mRNA molecule encoded by one of the inventive polynucleotides. Ribozymes possess highly specific endonuclease activity, which autocatalytically cleaves the mRNA.

Alternately, modulation may involve an increase in the expression and or activity of the polypeptide by over-expression of the polynucleotide, or by increasing the number of copies of the polynucleotide in the genome of the host.

To give effect to the above strategies, the invention also provides genetic constructs usually DNA constructs. The DNA constructs include the intended DNA (such as one or more copies of a polynucleotide sequence of the invention in a sense or anti-sense orientation or a polynucleotide encoding the appropriate ribozyme), a promoter sequence and a termination sequence (which control expression of the gene), operably linked to the DNA sequence to be transcribed. The promoter sequence is generally positioned at the 5' end of the DNA sequence to be transcribed, and is employed to initiate transcription of the DNA sequence. Promoter sequences are generally found in the 5' non-coding region of a gene but they may exist in introns (Luehrsen Mol. Gen. Genet 225, 81-93 (1991)) or in the coding region.

A variety of promoter sequences which may be usefully employed in the DNA constructs of the present invention are well known in the art. The promoter sequence, and also the termination sequence, may be endogenous to the target plant host or may be exogenous, provided the promoter and terminator are functional in the target host. For example, the promoter and termination sequences may be from other plant species, plant viruses, bacterial plasmids and the like. Preferably, promoter and termination sequences are those endogenously associated with the germacrene-D/*delta*-cadinene synthase genes.

Factors influencing the choice of promoter include the desired tissue specificity of the construct, and the timing of transcription and translation. For example, constitutive promoters, such as the 35S Cauliflower Mosaic Virus (CaMV 35S) promoter, will affect the transcription in all parts of the plant. Use of a tissue specific promoter will result in production of the desired sense or antisense RNA only in the tissue of interest. With DNA constructs employing inducible promoter sequences, the rate of RNA polymerase binding and initiation can be modulated by external stimuli, such as light, heat, anaerobic stress, alteration in nutrient conditions and the like. Temporally regulated promoters can be employed to effect modulation of the rate of RNA polymerase binding and initiation at a specific time during development of a transformed cell. Preferably, the original promoters from the gene in question, or promoters from a specific tissue-targeted gene in the organism to be transformed are used. Other examples of promoters which may be usefully employed in the present invention include, mannopine synthase (mas), octopine synthase (ocs) and those reviewed by Chua *et al.* Science 244, 174-181 (1989).

The termination sequence, which is located 3' to the DNA sequence to be transcribed, may come from the same gene as the promoter sequence or may be from a different gene. Many termination sequences known in the art may be usefully employed in the present invention, such as the 3' end of the *Agrobacterium tumefaciens* nopaline synthase gene. However, preferred termination sequences are those from the original gene or from the target species to be transformed.

The DNA constructs of the present invention may also contain a selection marker that is effective in cells, to allow for the detection of transformed cells containing the construct. Such markers, which are well known in the art, typically confer resistance to one or more toxins. One example of such a marker is the NPTII gene whose expression results in resistance to kanamycin or hygromycin, antibiotics which are usually toxic to plant cells at a moderate concentration. Alternatively, the presence of the desired construct in transformed cells can be determined by means of other techniques well known in the art, such as PCR or Southern blots.

Techniques for operatively linking the components of the inventive DNA constructs are well known in the art and include the use of synthetic linkers containing one or more restriction endonuclease sites. The DNA construct may be linked to a vector capable of replication in at least one host system, additional to the intended destination host system for example, *E. coli*, whereby after each manipulation the resulting construct can be sequenced and the correctness of the manipulation determined.

The DNA constructs of the present invention may be used to transform a variety of plants including agricultural, ornamental and horticultural plants. In a preferred embodiment, the DNA constructs are employed to transform kiwifruit, apple, banana, tomato, cotton, rose, olive and potato plants, carnation, mango, papaya, freesia, orchids, lisianthus, gerbera, grape. In a particularly preferred embodiment the plant is a carnation.

As discussed above, transformation of a plant with a DNA construct including an open reading frame comprising a polynucleotide sequence of the invention wherein the open reading frame is orientated in a sense direction can, in some cases, lead to a decrease in expression of the polypeptide by co-suppression. Transformation of the plant with a DNA construct comprising an open reading frame or a non-coding (untranslated) region of a gene in an anti-sense orientation will lead to a decrease in the expression of the polypeptide in the transformed plant.

It will also be appreciated that transformation of other non-plant hosts is feasible, including well known prokaryotic and eukaryotic cells such as bacteria (e.g. *E. coli*, *Agrobacterium, Lactobacillus*), fungi, insect, and animal cells is anticipated. This would enable production of recombinant polypeptides of the invention or variants thereof. The use of cell free systems (e.g. Roche Rapid Translation System) for production of recombinant proteins is also anticipated (Zubay Annu Rev Genet 7, 267-287 (1973)).

The polypeptides of the invention produced in any such hosts may be isolated and purified from same using well known techniques. The polypeptides may be used in cell-free systems for enzymic synthesis of germacrene-D and/or *delta*-cadinene and/or *alpha*-cubebene and/or *delta*-elemene and/or *alpha*-ylangene and/or *alpha*-copaene and/or *gamma*-elemene and/or germacrene-A and/or *gamma*-cadinene and/or selinadiene and/or *alpha*-muurolene and/or germacrene B.

Techniques for stably incorporating DNA constructs into the genome of target plants are well known in the art and include *Agrobacterium tumefaciens* mediated introduction, electroporation, protoplast fusion, injection into reproductive organs, injection into immature embryos, high velocity projectile introduction, floral dipping and the like. The choice of technique will depend upon the target plant to be transformed.

Once the cells are transformed, cells having the DNA construct incorporated into their genome may be selected by means of a marker, such as the kanamycin resistance marker discussed above. Transgenic cells may then be cultured in an appropriate medium to regenerate whole plants, using techniques well known in the art. In the case of protoplasts, the cell wall is allowed to reform under appropriate osmotic conditions. In the case of seeds or embryos, an appropriate germination or callus initiation medium is employed. For explants, an appropriate regeneration medium is used.

In addition to methods described above, several methods are well known in the art for transferring DNA constructs into a wide variety of plant species, including gymnosperms angiosperms, monocots and dicots.

The resulting transformed plants may be reproduced sexually or asexually, using methods well known in the art, to give successive generations of transgenic plants.

The nucleotide sequence information provided herein will also be useful in programs for identifying nucleic acid variants from, for example, other organisms or tissues, particularly plants, and for pre-selecting plants with mutations in germacrene-D/*delta*-cadinene synthase genes or their equivalents which render those plants useful in an accelerated breeding program to produce plants in which the content of germacrene-D or *delta*-cadinene and its derivatives is modulated. More particularly, the nucleotide sequence information provided herein may be used to design probes and primers for probing or amplification of germacrene-D/*delta*-cadinene synthase encoding polynucleotides. An oligonucleotide for use in probing or PCR may be about 30 or fewer nucleotides in length. Generally, specific primers are upwards of 14 nucleotides in length. For optimum specificity and cost effectiveness, primers of 16-24 nucleotides in length are preferred. Those skilled in the art are well versed in the design of primers for use in processes such as PCR.

If required, probing can be done with entire restriction fragments of the gene disclosed herein. Naturally, sequences based upon Figure 4 or the complements thereof can be used. Such probes and primers also form aspects of the present invention.

Methods to find variants of the of polynucleotides of the invention from any species, using the sequence information provided by the invention, include but are not limited to, screening of cDNA libraries, RT-PCR, screening of genomic libraries and computer aided searching of EST, cDNA and genomic databases. Such methods are well known to those skilled in the art.

The invention will now be illustrated with reference to the following non-limiting Examples.

EXAMPLES

Plant material and GC-MS analysis: Petals of *A. deliciosa* ([A. chev] C.F. Liang et A.R. Ferguson var. *deliciosa* 'Hayward') were taken at anthesis from vines grown in a HortResearch orchard at Te Puke New Zealand in 2000. Branches containing flowers were transported with stems in water, immediately after harvest. Whole female flowers (total weight 3.5g, at all stages of petal unfolding, and in good quality) were picked just under the receptacle and placed into a 250 mL Quickfit® Erlenmeyer flask to which was fitted a headspace adaptor with an air inlet and outlet. Volatiles were collected onto a Tenax-TA adsorbent trap, which was fitted to the air outlet port of the adaptor. The closed system was allowed to equilibrate for 30 min at room temperature (ca. 23 °C), after which the volatiles were purged from the vessel onto the trap with clean air at 20 mL min^{-1} for 100 minutes. The traps were stored at -15°C for up to two weeks, and then subjected to GC-FID/MS analysis.

Analysis of headspace samples: Volatile compounds were thermally desorbed from the headspace traps at 175 °C and were cryo-focussed at the beginning of the GC column (Young, 1981). The column outlet was split between the GC (Hewlett Packard 5890) FID detector (for quantitation) and VG-70SE (VG-Micromass, Manchester, U.K.) mass spectrometer (for component identification) with an electron impact ionisation of 70 eV. Separations were carried out in a 30 m x 0.32 mm i.d., 0.5 μm J & W DB Wax capillary column starting at 30 °C for 2 min, increasing by 3 °C min^{-1} to 50 °C, 5 °C min^{-1} to 130°C, to 240 °C and held for 2 min. Injector temperature was 240 °C. The carrier gas was He at a flow rate of 30 cm s^{-1}, and the FID and mass spectrometer transfer line were at 220°C. Quantification of compounds was carried out using an average detector response based on methyl butanoate, ethyl butanoate, hexanol and methyl benzoate. Components were identified by comparison with spectra in the mass spectral database (1998 NIST and in-house database), retention indices (in-house database) and in some cases direct GCMS comparison with authentic standards.

Isolation of mRNA and construction of cDNA library: Total RNA was extracted from *A. deliciosa* petals by an adaptation of the method of Gomez and Gomez (Langenkamper, *et al.*, 36, 857-869 (1998)). mRNA was purified from the total RNA by oligo(dT)-cellulose chromatography (Pharmacia) and was used to construct a Lambda ZAP-CMV (Stratagene) cDNA library according to the manufacturer's instructions. The cDNA-containing pBK-CMV plasmids were massed excised and used to transform *Escherichia coli* XLOLR (Stratagene). Plasmid was isolated from the XLOLR colonies and partially sequenced; the resulting cDNA sequences were then entered into the HortResearch EST database. All sequences on the database were compared to the NRDB90 database (ref) using the BLAST program (Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.). Putative terpene synthase cDNA sequences were identified by their similarity to known terpene synthases based on key protein motifs. A full-length sequence encoding a terpene synthase (EST75565) was identified. EST75565 was fully sequenced to determine that a functional cDNA suitable for *in vitro* expression studies had been isolated.

Computational analysis of sequence information: Computational analysis was performed using the European Molecular Biology Open Software Suite (EMBOSS). (Rice *et al*, 2000). The Transeq program was used to determine the amino acid sequence and the Pepstats program was used for the determination of molecular weight and isoelectric point. Sequence identity and similarity was calculated using the pair wise alignment program Needle, which uses the algorithm of Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)). The default parameters were used (Gap extension penalty: 0.5 for any sequence; Gap opening penalty: 10 for any sequence). Amino acid sequence alignments were performed using the program CLUSTALW (Thompson et al., 1994), and trimmed and shaded using the program GeneDoc (Nicholas and Nicholas, 1997).

Cloning into pGEM-T Easy and pET-30 vectors: For functional expression, a 1735 bp cDNA fragment encoding EST75565, excluding the initiating methionine, was amplified

by PCR using the following primers: forward primer: 5'- GAA TTC CAA CTA CCT TGT GCT CAA GC – 3'; and reverse primer: 5' – CTC GAG CCT CCA CTT CAG TGT CTT G – 3' (restriction sites underlined). PCR reactions were carried out in a total volume of 50 mL with the following reagents: 1x Expand High Fidelity PCR buffer (Boehringer), 0.2 mM dNTP's (Boehringer), 0.2 μM of each primer, and 1.75 units of Expand High Fidelity polymerase (Boehringer). PCR cycling was 94 °C (4 min); 25 cycles of 94 °C (30 s), 55 °C (30 s), 72 °C (2 min); and a final extension period at 72 °C (10 min). PCR products were purified using the QIAquick PCR clean up system (Qiagen) following the manufacture's specifications and cloned into pGEM-T Easy (Promega). A pGEM-T Easy clone harbouring a potential cDNA was fully sequenced to check for PCR errors, then excised using *Eco* RI and *Xho* I and subcloned into the expression vector pET-30a, yielding plasmid pET-30a75565. pET-30a75565 was resequenced at the 5' end to ensure the inserted cDNA was in frame, and then transformed into *E. coli* BL21-CodonPlus™-RIL cells (Stratagene).

Expression and characterization of germacrene/cadinene synthase from bacterial cultures: *E. coli* BL21-Plus™-RIL cells harbouring pET30a75565, or empty pET30a vector as a control, were grown overnight at 37 °C in Luria-Bertani media supplemented with 30 μg/mL kanamycin and 50 μg/mL chloramphenicol. A 500 mL aliquot of overnight culture was used to inoculate 50 mL of fresh 2 x YT medium supplemented with 30 μg/mL kanamycin and 50 μg/mL chloramphenicol. The culture was grown at 37 °C with vigorous agitation to A_{600} = 0.6 before induction with 0.3 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) and simultaneous addition of both farnesyl diphosphate (FDP) (100 μM) and hexadecane (required as an internal standard) or geranyl diphosphate (GDP) and hexadecane. The culture was immediately transferred to an incubator set at 16 °C, 30 °C or 37 °C depending on the experiment.

Headspace analysis of bacterial cultures: The headspace in the vessels above the bacterial cultures was collected immediately after the addition of FDP (or GDP) and hexadecane using solid phase micro extraction (SPME). The SPME fibres (65 μm PDMS/DVB, Supelco, Australia) were conditioned for 45 min at 260 °C and the

background analysed for contamination using GC-FID (HP5890) prior to use. The headspace volatiles were collected for 4 hours at 30°C with continuous agitation (110 rpm). Prior to analysis using a GC-FID/MS, the fibres were stored at ambient temperature in septum sealed glass vials. The volatiles were desorbed from the fibres for 5 minutes at 250°C in the GC injection port. The GC system was equipped with a DB-Wax capillary column (J & W Scientific, Folsom, USA), 30 m x 0.25 mm i.d., with a 0.5 μm film thickness. The carrier gas was helium at a flow rate of 30 cm s^{-1}. The GC oven was programmed to remain at 30°C for 6 min, then to increase by 3 °C min^{-1} to 102 °C, followed by an increase of 5 °C min^{-1} to 210 °C, which was maintained for 11 min. The mass spectrometer operated in electron impact ionisation (EI-MS) mode at 70 eV with a scan range of 30 – 320 amu. Peak identification was carried out by comparison of sample spectra with those from NIST, Wiley, and our own mass spectra libraries and was confirmed by retention indices of authentic standards and literature values (Davies 1998). Quantitative data was obtained by measuring sample peak area relative to hexadecane internal standard.

Characterization of germacrene/cadinene synthase from bacterial extracts and purified germacrene/cadinene synthase recombinant protein: Cultures were set up, grown and induced as above. Following induction, cultures were immediately transferred to a 16°C incubator and allowed to grow for a further 24 hours with continuous agitation and then harvested by centrifugation (2000 x g for 10min). Pelleted cells were resuspended in 20 mL binding buffer (5 mM imidazole, 0.5 mM NaCl, 10 mM DTT, 20 mM Tris-HCl (pH 7.9) Cells were disrupted with 2 times exposure to 12,700 psi in a French Pressure Cell Press (American Instrument Co. Inc, Silver Spring, Maryland. USA) and then centrifuged at 10000 x g for 15 min. 5 mL of supernatant was transferred to a 50 mL test-tube and adjusted to 10 mM $MgCl_2$ and 20 μM $MnCl_2$. FDP (or GDP) was added to a final concentration of 100 μM and the reaction mixture was incubated at 30 °C with shaking (110 rpm). Headspace volatiles were collected in the same manner as for whole cultures. The remainder of the extract (15 mL) was applied to a PD-10 gel filtration column (Amersham-Pharmacia Biotech) pre-equilibrated with binding buffer (DTT omitted). Eluent fractions were then pooled and purification of recombinant

protein was carried out in a single step using immobilised metal affinity chromatography (IMAC) utilising the recombinant proteins hexa-histidine (His_6 or His) tag. The eluent was applied to a Hi –Trap Chelating HP column (Amersham-Pharmacia Biotech) charged with Ni^+. Non-bound proteins were removed and recombinant protein was eluted following the manufacturer's specifications. 5 mL samples of the eluted protein were transferred to 50 mL test tubes and adjusted to 100 μM FDP (or GDP), 10 mM $MgCl_2$, 20 μM $MnCl_2$. Headspace volatiles were collected as for the bacterial cultures. A 5 mL purified sample was also trapped without the addition of any precursors as a further control. Aliquots of the remaining protein fraction were stored at -80°C in 20% glycerol until required for further analysis.

Electrophoresis and western analysis: Whole culture, His-purified and non His-purified protein extracts were analysed by SDS-PAGE. Protein bands were separated on a polyacrylamide gel comprising a 4% stacking layer and a 10% separation layer. Protein bands were either visualised using modified Neuhoff stain or were transferred on to Immobilin-P PVDF membrane (Millipore) using a Trans-Blot semi-dry electrophoretic transfer cell (Bio Rad). Blotted proteins were incubated with Anti-His_6 monoclonal antibody (Roche) and then with Anti-Mouse IgG-AP (Stressgen) secondary antibody, and were detected using 1-STEP™ NBT/BCIP (Pierce) alkaline phosphatase detection reagent.

Protein quantification: The protein concentration of crude and partially purified protein extracts were determined using the Bio Rad Protein-Assay™ reagent according to the manufacturers specifications. The reactions were quantified at 595 nm with a SPECTROmax PLUS spectrophotometer (Molecular Devices) using bovine serum albumin (BSA) as a standard.

Preparation of competent Agrobacterium tumefaciens GV3101: A. tumefaciens, strain GV3101, containing binary vectors for plant transformation, was inoculated into 10mL of LB or 2YT media containing rifampicin (10 mg/mL); gentamycin (25 mg/mL); and spectinomycin (100 mg/mL) for 24 hours at 28 °C, with shaking at 200 rpm. These

cultures were then used to inoculate a further 100-200mL of LB or 2YT media, with antibiotics as above, which were again grown for 24 hours at 28 °C, with shaking. The cells were collected by centrifugation (3,500 x g, 10 min, 4 °C) and resuspended, to a final OD600 of 0.8, in 5% sucrose solution. Silwet L-77 was added to a concentration of 0.05%.

Transformation of Agrobacterium tumefaciens GV3101: 45 mL aliquots of competent Agrobacterium cells were thawed gently on ice. 50-200 ng of plasmid DNA was added to each aliquot and gently mixed, then 40 mL of the cell/plasmid mixture was pipetted into a pre-chilled electroporation cuvette (0.2 cm gap, Bio-Rad). The cells were electroporated using a BioRad GenePulser, on the following settings:
Voltage: 2.5 kV
Capacitance: 25 mFd
Resistance: 400 Ohms
The time constant for the pulse was typically 7-9 ms.

The cells were immediately recovered by addition of 1 mL LB media, then decanted into sterile 15 mL centrifuge tubes and incubated at room temperature, with shaking (60 rpm). After 2 hours, 10 mL and 100 mL of the transformed bacteria was spread onto separate LB plates containing rifampicin (10 mg/mL); gentamycin (25 mg/mL); and spectinomycin (100 mg/mL); then grown for 48 hours at 28-30 °C.

Transformation of Arabidopsis thaliana by floral dip: Agrobacterium tumefaciens, containing the appropriate plasmids, were grown in 5mL cultures of LB media containing rifampicin (10 mg/mL); gentamycin (25 mg/mL); and spectinomycin (100 mg/mL) for 24 hours at 28 °C. The cells were collected by centrifugation and resuspended, to a final OD600 of 0.8, in 5% sucrose solution. Silwet L-77 was added to a concentration of 0.05%.

Healthy Arabidopsis plants, around five weeks old, showing a number of immature flower clusters, were used for the dipping procedure. The whole of the aboveground

portion of the plant was dipped into the Agrobacterium suspension, and gently agitated for 3-5 seconds. The dipped plants were then placed in humidity chambers in reduced light for 2-3 days, before being allowed to flower and set seed as normal. The seed was harvested upon complete drying of the plants (5-6 weeks after dipping).

Growth of seed from transformed Arabidopsis: Approximately 1000 T1 seeds were measured into a microcentrifuge tube. The seed was sterilised by a 1.5% bleach solution containing 0.01% Triton-X, mixed, and incubated for 15 min with occasional mixing. The seed was washed several times with distilled water, and resuspended in 0.1% agarose, prior to plating on 0.5X MS media, containing 100 mg/mL kanamycin. The plates were placed into growth rooms with a 12-hour light/12-hour dark cycle. After 2-3 weeks growth the plants were transferred onto fresh plates to confirm kanamycin resistance and were allowed to continue growing in glasshouse conditions until required for headspace analysis.

Headspace analysis of Arabidopsis: Plates containing the growing transformed and wild type Arabidopsis plants were transferred to appropriately sized gas jars fitted with an opening port for the GC-FID/MS fibres. Headspace volatiles were trapped and analysed as previously.

Results

Headspace analysis of volatiles emitted from kiwifruit flowers: Germacrene is a commonly found constituent of plants and exists as one of two isomers. (S)-(-)-germacrene [Fig 1] mainly occurs in higher plants, whereas (R)-(+)- germacrene [Fig 1] is predominantly found in lower plants such as liverworts (König *et al* Phytochem 43, 629-633 (1996)). In the headspace of the *A. deliciosa* petals analysed, only the (S)-(-)-germacrene isomer was detected, at a concentration of approximately 77.94 ng /g of petal tissue. Other sesquiterpenes included germacrene-B, α-farnesene, β-cubebene and one unknown sesquiterpene.

Sequence analysis of germacrene/cadinene synthase: Sequencing of EST75565 revealed an insert size of 1995 base pairs excluding the poly (A) tail. The cDNA sequence had an ORF encoding 565 amino acids with a putative start codon 82 bases downstream of the 5' end. The molecular mass of EST75565 was calculated to be 65 kDa. The predicted amino acid sequence of the germacrene/cadinene synthase does not contain a chloroplast-signalling peptide sequence (Emanuelsson, Nielsen, Brunak and von Heijne, 2000). These signalling peptides are typical of monoterpene and diterpene synthases and its absence is suggestive of it being a sesquiterpene synthase. As has been found for most other terpene synthases the predicted amino acid sequence of germacrene/cadinene synthase also contains a DDXX(D,E) motif (DDIYD) at amino acids 317 to 321 (Fig 4). This motif is involved in the binding of divalent metal ions necessary for catalysis. Germacrene/cadinene synthase was also shown to contain the angiosperm sequiterpene consensus sequence GVYXEP (GVYFEP) (Cai *et al* Phytochem 61, 523-529 (2002)), from amino acids 292 to 297.

Bohlmann, Meyer-Gauen and Croteau (Proc. Natl. Acad. Sci. USA 95, 4126-4133 (1998)) compared the amino acid sequences of 33 terpene synthases and showed that there were seven absolutely conserved amino acid residues. Our germacrene/cadinene synthase contains all seven of these absolutely conserved amino acids (Fig 4). They also found that six positions were absolutely conserved for aromatic amino acids and four positions were absolutely conserved for acidic amino acids. In our germacrene/cadinene synthase, four of the six aromatic positions and three of the four acidic positions are conserved.

The predicted isoelectric point for germacrene/cadinene synthase is 5.6 which is similar to the isoelectric point calculated for other sesquiterpene synthases. For example, two sesquiterpene synthases isolated from *Artemisia annua*, cASC34 and cASC125, have isoelectric points of 5.28 and 5.50, respectively (Van Geldre, *et al.*, Plant Sci. 158, 163-171 (2000)).

BLAST searches revealed that the predicted amino acid sequence of germacrene/cadinene synthase is most similar to the rose germacrene-D synthase (Guterman *et al,* Plant Cell 14, 2325-2338 (2002)), having 59 % identity and 74 %

similarity. Two sesquiterpene synthases from *Gossypium arboreum* (tree cotton), *delta*-cadinene synthase (Liang, Genbank, 1998), and *delta*-cadinene synthase isozyme A (Chen *et al.*, J. Nat. Prod. 59 (10), 944-951 (1996)) with only two base pairs of difference, show the second highest similarity having 52% identity and 69% similarity at the amino acid level for both proteins. A sesquiterpene synthase from *Lycopersicon hirsutum* (tomato) of unknown function has the third highest similarity (van Der Hoeven *et al.*, Plant Cell 12 (11), 2283-2294 (2000)), with 49% identity and 69% similarity. Specific regions of amino acid identity between these synthases are generally short (usually between 5 and 10 amino acids). The most notable homologies between all four proteins is a region immediately upstream of and including the putative angiosperm sesquiterpene synthase consensus sequence GVYFEP (Cai *et al* Phytochem 61, 523-529 2002). This includes a region of 13 or 14 identical amino acids for rose germacrene D synthase and the *Lycopersicon hirsutum* sesquiterpene synthase. The rose germacrene-D synthase also shares a region of 16 identical amino acids with EST 75565 from amino acids 314 to 330. Other regions of identity between these synthases include 4 amino acids from amino acid 167 to 170 and 4 amino acids from amino acid 265 to 268.

Western analysis: Western analysis confirmed the presence of a soluble expression product that was slightly larger (ca. 70 kDa) than the predicted 65kDa for the native protein. This size difference can be attributed to a His_6 tag being present in the recombinant EST75565 expression product. This product was detected in both the bacterial and partially purified recombinant protein extracts. No equivalent bands were detected in any of the vector only control extracts.

Characterisation of germacrene/cadinene synthase:

Germacrene-D (80%), a lesser amount of cadinene, primarily *delta* isomer (20% and small quantities (2%) of a range of sesquiterpenes - humulene, selinene, γ-cadinene, germacrene-B muurolene, amorphene, ylangene, zingiberene and aromadendrene were detected in the headspace of bacterial cultures harbouring pET-30a75565 (Fig 6b). In addition isomers of elemene and elemol were found, which are thermal rearrangements products of some germacrene isomers (A, B and C but not D). Thermal rearrangement

occurs at the high temperatures used for SPME desorption. Controls comprising *E. coli* BL21 cells transformed with pET-30a lacking the germacrene-D/*delta*-cadinene synthase cDNA insert produced none of these compounds (Fig 6a). Addition of geranyl diphosphate (GDP) to bacterial cultures did not result in the production of any monoterpenes but still produced germacrene-D, although at significantly lower levels to that exhibited with FDP addition. This is due presumably to the presence of bacterial FDP. Headspace analysis of partially purified recombinant enzyme also produced the germacrene-D and B, delta- and *gamma*-cadinene, elemene isomers and elemol. However, other products were also apparent, including germacrene-A, copaene and an unidentified sesquiterpene. (Fig 7). Germacrene-A and the unidentified sesquiterpene became major components of the headspace volatiles. The ratios present are *alpha*-cubebene (0.23%), *delta*-elemene (1.1%), *alpha*-ylangene (0.6%), *alpha*-copaene (0.15%), *gamma*-elemene (5.48%), germacrene D (19.57%), germacrene A (32.31%), *delta*-cadinene (2.48%), *gamma*-cadinene (2.26%), selinadiene (0.54%), *alpha*-muurolene (0.42%), germacrene-B (3.61%, elemol (3.39%) and an unknown sesquiterpene at r/t 53.53 (27.87%). These compounds were all shown to be dependent upon the addition of FDP. Addition of GDP or absence of any precursor to purified extracts did not result in the production of either monoterpenes or sesquiterpene compounds.

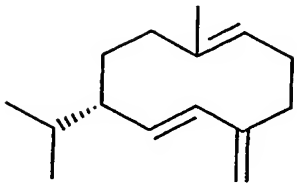
Expected outcome from *Arabidopsis* transformants:

Arabidopsis plants harbouring the EST75565 cDNA insert are expected to demonstrate an altered volatile profile, with altered production of germacrene-D, *delta*-cadinene and other sesquiterpene compounds when compared to wild type controls.

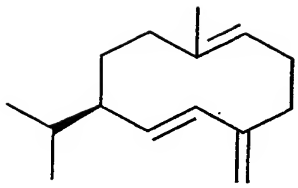
The above Examples are an illustration of practice of the invention. It will be appreciated by those skilled in the art that the invention can be carried out with numerous modifications and variations. For example, variations to the nucleotide sequences may be used and the sequences may be expressed in different organisms.

By the authorised agents
A J PARK
Per

Figure 1 Germacrene D enantiomers

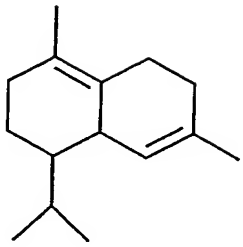


R-(+)-Germacrene D



S-(-)-Germacrene D

Figure 2 Structure of δ-cadinene



δ-cadinene

Figure 3 Pathway showing formation of germacrene D and δ-cadinene from FDP

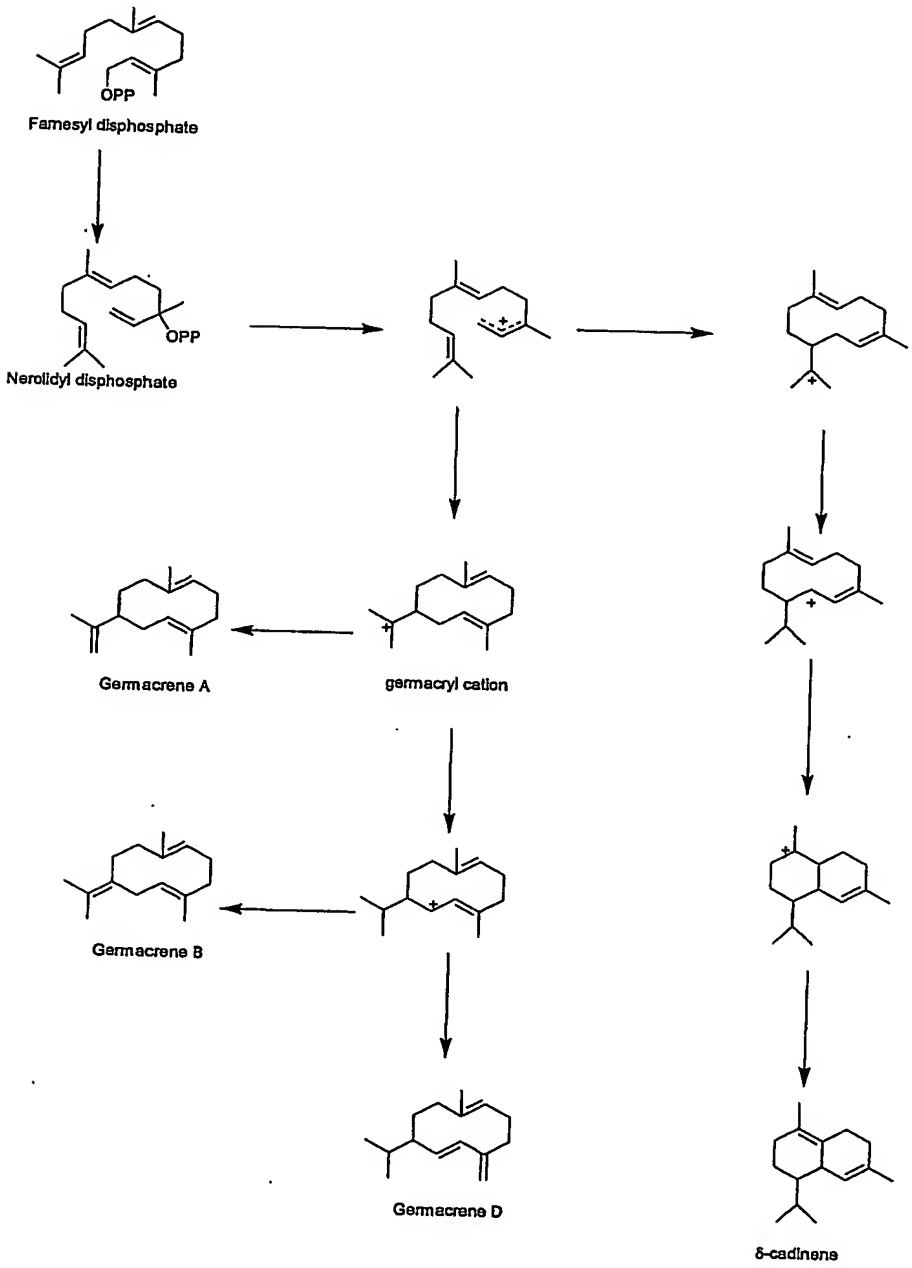


Figure 4 Polynucleotide sequence

```
1    GTGAAAACTA AAATAGGCCA AGTGTGTAGG TTCATCTCTA GTTTTTCTCT TTAAATTAAT TCTTCAACCC AGAAAAAAAA CATGCAACTA CCTTGTGCTC
101  AAGCTTTGCC AATACCAACT GTTACAACCA CCACTAGTAT TGAACCAGCA CATGTAAGGC GTCGCTCTGC AAATTATCAT CCTAGCATTT GGGGAGATCA
201  TTTCCTCGCC TACTCTTCCG ATGCTATGGA AGAAGAGGTT ATTAACATGG AACAACAACA ACGACTTCAT CACCTGAAAC AAAAGGTGAG AAAAATGCTA
301  GAGGCAGCTG CTGAACAATC TTCACAGATG CTGAACCTGG TCGACAAAAT CCAACGCTTA GGCGTGTCTT ACCATTTTGA AACTGAGATC GAAACAGCTT
401  TACGGCACAT ATACAAAACC TGTGATTACC ATTTTGATGA TCTCCACACT GCTGCTCTCT GTTTTCGGTT ACTTAGACAA CAAGGATATC CAGTTTCTTG
501  TGATATGTTC GACAAATTCA AGAACAGCAA AGGTGAGTTT CAAGAATCCA TAATCAGCGA TGTGCAAGGA ATGTTAAGTT TGTATGAAGC TACATGTCTA
601  AGGATACACG GAGAAGATAT ACTAGACGAA GCACTAGCTT TTACCATCAC TCAACTTCGG TCCGCATTGC CCAACTTAAG CACTCCTTTC AAGGAACAAA
701  TCATTCATGC TCTGAACCAG CCCATCCACA AGGGGTTGAC AAGGCTCAAG GCAAGGAGCC ACATTTTATT TTTTGAACAG AATGATTGCC ATAGCAAAGA
801  CCTTTTGAAT TTCGCAAAAT TAGATTTCAA CTTATTACAA AAGTTGCACC AGAGGGAGCT ATGTGAAATC ACAAGGTGGT GGAAAGATTT GAATTTTGCA
901  AAGACACTAC CTTTTGCCAG AGACAGAATG GTAGAGTGCT ACTTTTGGAT ACTTGGGGTG TACTTTGAGC CCCAATATCT GCTTGCTAGG AGGATGCTAA
1001 CCAAGGTGAT TGCCATGATT TCCATTATCG ATGACATCTA CGATGTCTAC GGTACCTTGG AAGAACTTGT TCTCTTCACT GATGCAATTG AGAGGTGGGA
1101 GATCAGTGCC TTGGATCAAC TTCCAGAGTA TATGAAACTA TGTTATCAAG CACTTTTGGA TGTTTATAGT ATGATTGATG AAGAGATGGC GAAGCAAGGA
1201 AGATCTTATT GCGTAGACTA TGCAAAATCT TCAATGAAAA TTTTGGTTAG AGCATACTTC GAAGAAGCCA AATGGTTTCA CCAAGGATAT GTTCCAACTA
1301 TGGAAGAGTA TATGCAAGTT GCATTAGTAA CCGCGGGTTA CAAAATGCTT GCAACCTCTT CCTTTGTTGG CATGGGAGAT TTGGCAACCA AAGAGCCCTT
1401 TGATTGGGTG TCAAATGATC CTTTAATTGT TCAAGCTGCA TCAGTGATAG GCAGACTCAA GGATGACATT GTTGGCCACA AGTTTGAGCA AAAGAGAGGG
1501 CACGTGGCGT CGGCTGTGGA ATGCTACAGT AAGCAACATG GTACAACAGA GGAAGAGGCT ATTATTGAAT TGGATAAACA AGTTACACAT TCATGGAAAG
1601 ACATCAAGGC AGAGTGCCTC TGCCCAATCA AGGTCCCAAT GCGTCTTCTT GCGCGAGTTC TCAATCTTGC ACGAGTGCTT TATGTTATAT ACCAGGATGA
1701 AGACGGATAC ACTCATGCTG GAACCAAGGT CGAGAACTTT GTAACCTCAG TGCTTATCGA TTCTATGCCA ATCAATTAGA AAATGTAACA AGACACTGAA
1801 GTGGAGGCAT AAATAAATTC AAAAGTTGAT TTAAAGTTGG GGTAGTGAAC GGGGATTCTT ACCATTAAGA GATATTCTTG CTAAAAAGCA ATTAATTCAA
1901 TGCATTTCCA ATAAAATAAT TTAGCCAGTT GTTCTTCATC TTGTTTTTTT TTTGTTTCTC TTTCCTTTCT AAATATAAAA TTATAATTAA TTGGCAAAAA
2001 AAAAAAAAAA AAAAAAAAA
```

Figure 5 Polypeptide sequence

```
1   MQLPCAQALP IPTVTTTTSI EPPHVTRRSA NYHPSIWGDH FLAYSSDAME
51  EEVINMEQQQ RLHHLKQKVR KMLEAAAEQS SQMLNLVDKI QRLGVSYHFE
101 TEIETALRHI YKTCDYHFDD LHTAALSFRL LRQQGYPVSC DMFDKFKNSK
151 GEFQESIISD VQGMLSLYEA TCLRIHGEDI LDEALAFTIT QLRSALPNLS
201 TPFKEQIIHA LNQPIHKGLT RLNARSHILF FEQNDCHSKD LLNFAKLDFN
251 LLQKLHQREL CEITRWWKDL NFAKTLPFAR DRMVECYFWI LGVYFEPQYL
301 LARRMLTKVI AMISIIDDIY DVYGTLEELV LFTDAIERWE ISALDQLPEY
351 MKLCYQALLD VYSMIDEEMA KQGRSYCVDY AKSSMKILVR AYFEEAKWFH
401 QGYVPTMEEY MQVALVTAGY KMLATSSFVG MGDLATKEAF DWVSNDPLIV
451 QAASVIGRLK DDIVGHKFEQ KRGHVASAVE CYSKQHGTTE EEAIIELDKQ
501 VTHSWKDINA ECLCPIKVPM PLLARVLNLA RVLYVIYQDE DGYTHPGTKV
551 ENFVTSVLID SMPIN*
```

Figure 6. Whole cultures a) Pet30a control + FDP and b) Pet30a/75565 + FDP

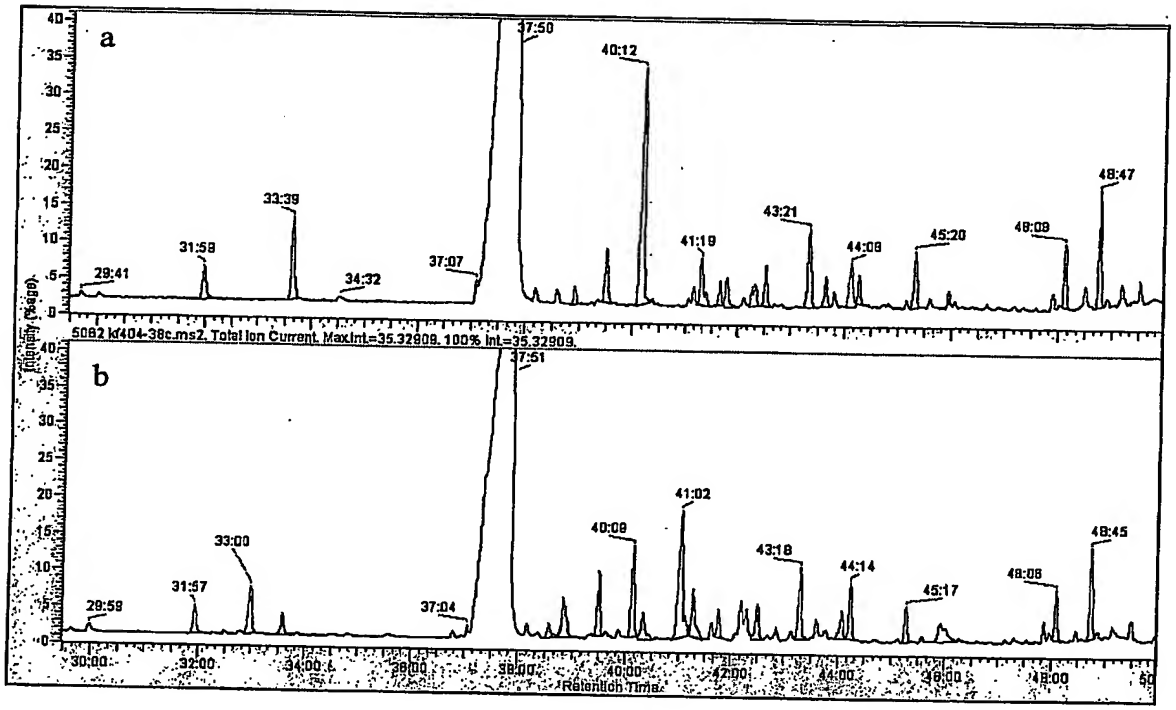


Figure 7. Purified extracts a) Pet30a control + FDP and b) Pet30a/75565 + FDP